OA-17

Progressive prediction using instrumental variable for accurate prediction of band gap of ABO_3 perovskites

Ch. Li[1], H. Hao[1], G. Zhao[1], H. Liu[1*], B. Xu[2*]

[1]*State Key Laboratory of Advanced Technology for Materials Synthesis and Processing and International School of Materials Science and Engineering, Wuhan University of Technology, Wuhan 430070, P.R. China*

[2]*State Key Laboratory of New Ceramics and Fine Processing School of Materials Science and Engineering, Tsinghua University, Beijing 100084, P.R. China*

[*]*lhxhp@whut.edu.cn; xuben@mail.tsinghua.edu.cn*

Perovskite-type oxide materials have been extensively used in many technologically applications due to their unique physical and chemical properties [1, 2]. The band gap is a key parameter that governs the conductivity of superconductor and the efficiency of perovskite solar cells. And it is usually measured by experiments or quantum mechanics calculations, but the drawbacks of expensive and time-consuming make the screening of possible compounds impractical [3, 4]. Therefore, a rapid and accurate method for predicting band gap of perovskites is urgently required. In this work, we proposed a robust machine-learning framework with instrumental variable – formation energy – to predict the band gap of perovskites progressively. An optimal feature set containing 24 attributes was established, and the modulus of bond-valence vector sum [5] on three sites (A, B and O site) were chosen to represent the structural distortion. After establishing a robust formation energy regression model (R^2:0.964), the predicted value was used as an instrumental variable to predict the band gap, and a tremendous precision (R^2:0.857) was obtained. A structure-property relationship mapping perovskites band gap is concurrently excavated that the number of electrons in *d* orbital, formation energy and modulus of bond-valence vector sum on O site are the three most crucial and relevant predictors.

Figure 1 shows the overall workflow of the progressive learning method. Schematic presents the details of perovskite compunds collection and the outlines of progressive learning workflow, including E_f prediction, instrumental variable generation, E_g prediction and results analysis.

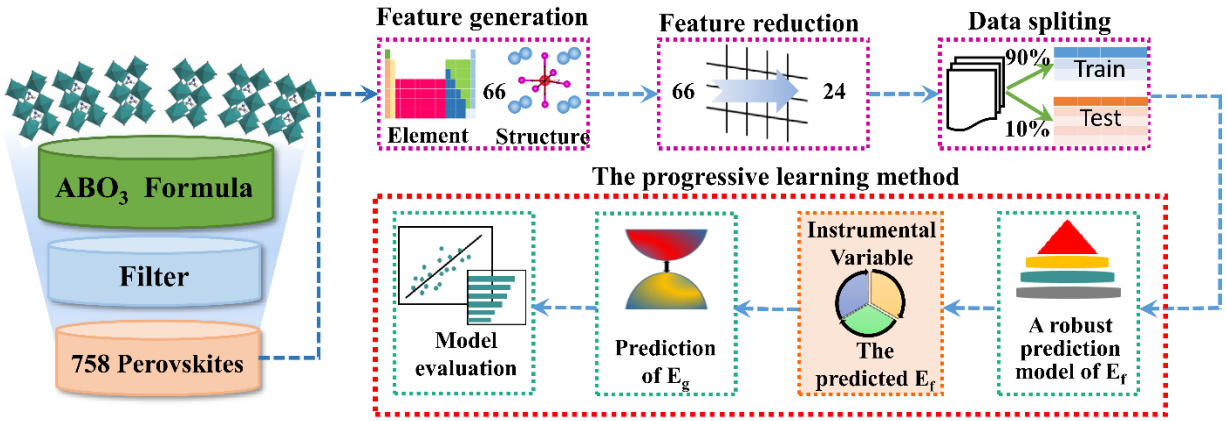

Figure 1. Overall workflow of the progressive learning method.

1. G. Pilania, P.V. Balachandran, J.E Gubernatis, T. Lookman, *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **71**, 507 (2015).
2. C. Ablitt, S. Craddock, M.S. Senn, A.A. Mostofi, N.C. Bristowe, *npj Computational Materials* **3**, 44 (2017).
3. P. Geerlings, F.D. Proft, W. Langenaeker, *Chemical Reviews* **103**, 1793 (2003).
4. E. Kim, K. Huang, S. Jegelka, E. Olivetti, *npj Computational Materials* **3**, 53 (2017).
5. Y. Qi, S. Liu, I. Grinberg, A.M. Rappe, *Physical Review B* **94**, 134308 (2016).